Chapter 1

MIMO Systems: Multiple Antenna Techniques

Yiqing ZHOU, Zhengang PAN, Kai-Kit WONG

1Dr, Senior Member of IEEE, Associate Editor, IEEE TWireless, IEEE CL, and JoC (AP), Senior Lecturer, Department of EEE, University College London, London, WC1E 7JE, United Kingdom
Tel: +44-(0)1473-663724, Fax: +44-(0)1473-635199
E-mail: yqzhou@astri.org, zgpan@astri.org, k.wong@ee.ucl.ac.uk
1.1. Abstract

This chapter reviews most known multiple antenna techniques for single-user point-to-point systems, from how multiple antennas help provide diversity and multiplexing to the detection techniques for these systems. Finally, this chapter also discusses the possibility of using multiple antennas for multiuser systems for spatial multiplexing.

1.2. Literature Review

The concept of multiple antennas can be traced back long time ago. In 1960’s, antenna array was applied in military radar systems for signal copying, direction finding and signal separation [1,2]. These signal parameter estimations need high resolution. Many algorithms have been proposed in this field, such as the maximum likelihood (ML) based approach [3] and maximum entropy (ME) based approach [4]. The main limitations of these approaches are the bias and high sensitivity. These problems were solved by Schimidt [5,6] and Bienvenu [7] independently. Particularly, Schimidt's method is known as MUSIC (Multiple Signal Classification), which can provide substantial performance improvement at the cost of high computational complexity and storage consumption. Then, another scheme, ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques), is proposed to achieve performance close to that of MUSIC but with a significantly reduced complexity [8].

In wireless communication systems, due to the limited physical size of mobile devices, the application of multiple antennas is initially proposed for base stations (BS) only. Investigations on these multiple antenna systems focus on beamforming, estimation of direction of arrival (DOA), and spatial diversity. The aforementioned algorithms like MUSIC and ESPRIT can be employed to realize the beamforming and the estimation of DOA, where the received/transmitted signals from/different antennas are coherently combined to point at a specific direction. In this way, the co-channel interference from other mobile stations (MS) can be reduced in the uplink and the transmission power can be focused to the desired MS in the downlink. For example, in [9,10,11] and some references therein, multiple antennas are deployed at BS to obtain receive diversity and/or co-channel interference rejection in the uplink based on the DOA. It is demonstrated that significant capacity improvement can be achieved. Then, in [10,12], the antennas at the BS are used as a transmit beamformer in downlink. By focusing the transmit energy in the direction of the desired MS, transmit beamforming increases the signal-to-noise ratio (SNR) at the MS. Similar to the single-input single-output (SISO) system with only one antenna at both sides, the aforementioned receive diversity and transmit diversity achieving systems are called single-input multiple-output (SIMO) and multiple-input single-output (MISO) systems, respectively.

Nowadays, multiple-input multiple-output (MIMO) systems have become one of the hottest research topics due to the magnificent enhancement brought by multiple antenna techniques. The possibility to deploy multiple antennas at both sides of the communication link lays on the following facts: 1) the developments in hardware miniaturization and advances in antenna design make the deployment of multiple antennas at the small MS more feasible; 2) wireless applications like wireless local area network (WLAN) and fixed wireless access (FWA) allow large physical sized devices that can afford multiple antennas, such as laptop computers. The research on MIMO systems starts from early 1990’s. Telatar [13] found that using MIMO system, there could be a dramatic improvement in the system throughput when no extra spectrum is needed. In 1996, Foschini published the famous paper [14], where he proposed a Bell Labs Layered Space-Time (BLAST) architecture for MIMO systems and it was shown that high spectral efficiencies such as 10 to 20 bits/s/Hz can be achieved. Then, an elegant space time block coding (STBC) architecture was proposed by Alamouti [15], which is simple yet effective in obtaining spatial diversity. Alamouti code has led to a research fervor on STBC. In summary, MIMO can be employed to achieve both transmit and receive diversity, or make the system throughput increase linearly with the minimum number of the transmit and receive antennas.

1.3. Space-Time Coding

Transmitter diversity was firstly proposed in [16,17] for BS simulcasting, where copies of the same symbol are transmitted through multiple antennas at different time, creating an artificial multipath distortion at the receiver.
A ML sequence detection or a minimum mean square error (MMSE) equalizer can be employed to resolve the multipath interference and obtain diversity gain. This delay coding scheme can be taken as a repetition channel coding in spatial domain. The idea of channel coding in space and time domains was generalized by Tarokh et al. in [18] and the so-called space-time coding (STC) was introduced. There are several kinds of STC with different structures.

### 1.3.1. Space-Time Block Coding

Alamouti discovered an elegant STBC scheme for MIMO systems with two transmit antennas. As shown in Figure 1.1, the input symbols to the STBC encoders are divided into groups of two symbols. In the first symbol duration, two symbols, $s_1$ and $s_2$, are transmitted simultaneously from antenna #1 and #2, respectively. Then, in the next symbol duration, the signals $-s_2^*$ and $s_1^*$ are transmitted from antenna #1 and #2, respectively. Assuming that $h_{1,1}$ and $h_{1,2}$ are the fading coefficients corresponding to the first and second transmit antenna to the receive antenna, respectively, and are constant over two consecutive symbol time, the signals received over two consecutive symbol periods are

$$y_1 = \begin{pmatrix} h_{1,1} \\ h_{1,2} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + n_1 = h_{1,1}s_1 + h_{1,2}s_2 + n_1$$  \hspace{1cm} (1.1)

and

$$y_2 = \begin{pmatrix} h_{2,1} \\ h_{2,2} \end{pmatrix} \begin{pmatrix} -s_2^* \\ s_1^* \end{pmatrix} = -h_{2,1}s_2^* + h_{2,2}s_1^* + n_2$$ \hspace{1cm} (1.2)

They can be rearranged into a matrix form as follows

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & -h_{2,2} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix}$$ \hspace{1cm} (1.3)

where $\mathbf{H}_{\text{eff}}$ is the effective channel matrix. Since $\mathbf{H}_{\text{eff}}$ is orthogonal, by a simple linear operation $\mathbf{H}_{\text{eff}}^H \mathbf{y}$, the 2 by 1 MISO system can be split into two separate parallel SISO systems as follows

$$\mathbf{z} = \mathbf{H}_{\text{eff}}^H \mathbf{y} = \begin{pmatrix} h_{1,1} \\ -h_{1,2} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \begin{pmatrix} h_{1,1}n_1 + h_{1,2}n_2^* \\ -h_{1,1}n_1 + h_{1,2}n_2^* \end{pmatrix}$$ \hspace{1cm} (1.4)

![Figure 1.1. Alamouti code](image-url)
This orthogonal property can be extended to the case where the receiver has multiple antennas. It can be seen that the Alamouti scheme decouples the vector ML decoding problems into scalar problems. Therefore, the scheme can realize full diversity while reduce the receiver complexity dramatically. It will be shown later that the Alamouti code provides a performance gain similar to that obtained by using one transmit antenna and two receive antennas with maximum ratio combiner (MRC) except for a power reduction of 2 (3 dB).

The Alamouti scheme can be generalized to an arbitrary number of antennas [19] and is able to achieve the full diversity promised by the number of transmit and receive antennas. Note that in the previous example with two transmit antennas, two different data symbols are transmitted in two symbol durations. Thus the coding rate is $2$ symbols/2 durations $= 1$, or the full rate. However, for arbitrary number of antennas, full rate is not always achievable. Letting the $i^{th}$ row contain the data symbols transmitted at the $i^{th}$ antenna in different symbol durations, two STBC with coding rate $3/4$ and $1/2$ are given by

$$
\begin{pmatrix}
  s_1 & -s_2^* & s_3^*/\sqrt{2} & s_4^*/\sqrt{2} \\
  s_2 & s_1^* & s_3^*/\sqrt{2} & -s_4^*/\sqrt{2} \\
  s_3^*/\sqrt{2} & s_1^*/\sqrt{2} & -s_1-s_2+s_2^* & s_2+s_2^*+s_1-s_1^*
\end{pmatrix}
\quad (1.5)
$$

and

$$
\begin{pmatrix}
  s_1 & -s_2 & -s_3 & -s_4 & s_1^* & -s_2^* & -s_3^* & -s_4^* \\
  s_2 & s_1 & s_4 & -s_3 & s_2^* & s_1^* & s_4^* & -s_3^* \\
  s_3 & -s_4 & s_1 & s_2 & s_3 & -s_4^* & s_1^* & s_2^* \\
  s_4 & s_3 & -s_2 & s_1 & s_4 & s_3^* & -s_2^* & s_1^*
\end{pmatrix}
\quad (1.6)
$$

respectively.

**1.3.2. Space-Time Trellis Codes (STTC)**
Although STBC can achieve a near optimal diversity gain with a very simple decoding algorithm, its coding gain is limited. Tarokh [18] developed another STC approach known as space time trellis code (STTC) to achieve both diversity and coding gains. In STTC, symbols are encoded according to the number of antennas at which they are simultaneously transmitted and decoded using a ML sequence detector at the receiver. Given the number of transmit antennas, the code design of STTC aims to construct the largest possible codebook with diversity and coding gain. The design criteria in different scenarios have been developed by Tarokh et al., and a number of hand-crafted STTCs with full diversity gain were given [18].

Figure 1.2 shows the encoder and decoder of a 4-state STTC with quadrature phase shift keying (QPSK) modulation and two transmit antennas. As shown in the figure, there are 4 states of 0, 1, 2 and 3. The initial state is 0. If the input symbol is 2, the output will be two symbols, i.e., 0 and 2, and the state is transformed to 2. The two output symbols are transmitted from antenna #1 and #2, respectively. Next, a second data symbol 1 is input, then the output will be 1 and 2, and the state becomes 1. Using the trellis structure, at the receiver, a ML sequence detection can be carried out. In the STTC scheme, diversity gain is achieved since the encoded data arrives over uncorrelated faded branches. Moreover, STTC can provide coding gain because of the trellis structure which creates a code relationship in the time and space domains. These space-time trellis structures can be realized by shift registers with generator coefficients determining the multiple output symbols that are fed to different transmit antennas. STTC performs better than STBC at the cost of additional encoding/decoding complexity.

1.3.3. Space-Time Turbo/LDPC Codes

ST-Turbo code and ST-LDPC combine the turbo and low density parity check (LDPC) codes with space-time coding. Unlike the classical STC codes with designing rules like STTC and STBC, there lacks of systematic ways to design good ST-Turbo code and ST-LDPC. Most researches try to extend the Turbo-code and LDPC designed for traditional SISO systems to MIMO systems. Although limited analysis has been done, ST-Turbo and ST-LDPC have shown attractive system performance due to the powerful iterative decoding method. Figure 1.3 illustrates two typical encoder structures of ST-Turbo code, one parallel concatenated structure and one serial concatenated structure. In the parallel structure, STTC can be employed as the constituent codes, while the serial ST-Turbo is a concatenation of block/convolutional codes with STTC/STBC. Moreover, Figure 1.4 depicts a typical encoder structure of ST-LDPC which is a serial concatenation of one conventional LDPC and one STTC/STBC. ST-Turbo/LDPC codes can achieve both coding gain and spatial diversity gain. However, as shown in Figure 1.3, for parallel structured ST-Turbo codes, the number of transmit antennas are limited by the turbo code structure. At the receive side, for parallel ST-Turbo, STTC decoding is needed, while for the serial structure, conventional serial turbo decoding algorithm can be employed. Details of these code schemes can be found in [22] and [23] and the references therein.

1.3.4. Differential STC

The coherent decoding algorithm of above coding schemes requires the channel state information (CSI) at the receive side. However, in many cases, it is difficult to acquire the CSI, for example, in case of high fading rate or low complexity receiver structure. In these cases, non-coherent decoding would be a good choice. To this end, a
new class of differential STC is proposed in [24], an example of which is shown in Figure 1.5 with two transmit antennas. At time $T+2$, the input symbols vector $\mathbf{s}_{T+2}$ is differentially processed by the matrix $(\mathbf{C})_r$, which is generated from the STBC encoder output at time $T$. The resultant symbols are denoted as $\mathbf{c}_{T+2}$, given by

$$
\mathbf{c}_{T+2} = \begin{pmatrix} c_{1,T+2} \\ c_{2,T+2} \end{pmatrix} = (\mathbf{C})_r \mathbf{s}_{T+2} = \begin{pmatrix} c_{1,T} & -c_{2,T} \\ c_{2,T} & c_{1,T} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix}_{T+2}
$$

(1.7)
Assuming one receive antenna and ignoring the background noise, the received signals at time $T$ and $T+1$ can be written as

$$
y_T = \begin{pmatrix} y_T \\ y_{T+1} \end{pmatrix} = \begin{pmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & -h_{3,1} \end{pmatrix} \begin{pmatrix} c_{1,T} \\ c_{2,T} \end{pmatrix}
$$

Similarly, the received signals at time $T+2$ and $T+3$ are given by

$$
y_{T+2} = \begin{pmatrix} y_{T+2} \\ y_{T+3} \end{pmatrix} = \begin{pmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & -h_{3,1} \end{pmatrix} \begin{pmatrix} c_{1,T+2} \\ c_{2,T+2} \end{pmatrix} = \begin{pmatrix} h_{1,1}^* & h_{1,2}^* \\ h_{2,1}^* & -h_{3,1}^* \end{pmatrix} \begin{pmatrix} c_{1,T}^* \\ c_{2,T}^* \end{pmatrix} \begin{pmatrix} s_i \\ \end{pmatrix}_{T+2}
$$

Therefore, the signal transmitted at $T+2$ can be simply recovered by

$$
\begin{pmatrix} s_1 \\ s_2 \end{pmatrix}_{T+2} = \frac{1}{|y_T|^2 + |y_{T+1}|^2} \begin{pmatrix} y_T \\ y_{T+1} \end{pmatrix}^H \begin{pmatrix} y_{T+2} \\ y_{T+3} \end{pmatrix}
$$

1.4. SIMO

As shown in the previous sub-section, multiple transmit antennas are required to apply space-time coding. When there is only one antenna at the transmit side, SIMO can be used to achieve receive diversity. The capacity of SIMO system increases logarithmically to the number of antenna. An illustration of SIMO system is shown in Figure 1.6 with $n_r$ receive antennas, where a signal $s$ is transmitted from one antenna. After passing through the multiple antenna channel $h = (h_{1,1}, h_{2,1}, \ldots, h_{n_r,1})^T$, the received signals can be written into a vector as $y = (y_1, \ldots, y_{n_r})^T$, where $h_{i,j}$ stands for the channel factor from the transmit antenna to the receive antenna #i, and $y_i$ is the received signal at antenna #i ($i=1, 2, \ldots, n_r$). The relationship between $y$ and $h$ is given by

$$
y = h \cdot s + n
$$

where $n = (n_1, \ldots, n_{n_r})^T$ is the noise vector with a covariance matrix of $\sigma^2 \cdot I_{n_r \times n_r}$. According to the electromagnetic property of antennas, the antennas should be sufficiently spaced so that the channel fading at different antennas is independent to each other. Assume that the CSI is known at the receiver, which can be realized by channel estimation. In the following, two algorithms are introduced to obtain receive diversity.

![Figure 1.6. A SIMO system](image-url)
1.4.1. Selection Combining

At the receive side, the instant signal to noise ratio at antenna \( i \) is given by

\[
\text{SNR}_i = \frac{E\left|h_{i,i}^*s\right|^2}{E\left|h_i^2\right|^2} = \frac{|h_{i,i}|^2 P}{\sigma^2}
\]

(1.12)

where \( P = E\left|s^2\right| \) is the signal power. Since the channels experienced by receive antennas are different, the receive SNR varies from antenna to antenna. Using selection combining, the received signal at the antenna with the largest SNR is chosen as the output.

1.4.2. Maximum Ratio Combining

Different to selection combining, whose output is just one of the \( n \) received signals, maximum ratio combining (MRC) makes use of all received signals. The basic idea is that the received signal with higher SNR should have a more important role in the final signal. Therefore, the weighting vector of gain combining is given by

\[
w = (h_{i,1}, h_{i,2}, \ldots, h_{i,n}) = h^H
\]

(1.13)

After gain combining, the signal becomes

\[
y = w \cdot r = h^H (h \cdot s + n) = \|h\|^2 \cdot s + h^H \cdot n
\]

(1.14)

where \( \| \| \) stands for the Euclidean norm of a vector and \( \|h\|^2 = \sqrt{\sum_{i=1}^{n} |h_{i,i}|^2} \). The SNR at the output of gain combining is then given by

\[
\text{SNR} = \frac{E\left|h^H \cdot s\right|^2}{E\left|h^H \cdot n\right|^2} = \frac{\|h\|^2 \cdot P}{\|h\|^2 \cdot \sigma^2} = \frac{P}{\sigma^2} \sum_{i=1}^{n} |h_{i,i}|^2
\]

(1.15)

Compared to selection combining, MRC can produce a signal with higher SNR since \( \|h\|^2 = \sum_{i=1}^{n} |h_{i,i}|^2 \geq |h_{i,i}|^2 \) for any \( i \in \{1, n\} \). Moreover, a receive diversity of order \( n \) can be achieved.

Figure 1.7. A MISO system
1.5. MISO

The structure of a MISO system is shown in Figure 1.7, where the same data will be radiated from \( n_t \) transmit antennas. Similar to SIMO systems, the transmit antennas should be sufficiently spaced to ensure that the channel fading at different antennas is independent to each other. In SIMO systems, channel information is estimated at the receiver. Thus, selection combining and MRC can be employed to achieve receive diversity. But in MISO systems, there is no channel information at the transmitter. There are two different ways to obtain transmit diversity in MISO systems. If channel information is fed back from the receiver, precoding schemes like transmit MRC can be employed. On the other hand, if there is no channel information at the transmitter, preprocessing algorithms can be employed, such as space time coding.

1.5.1. Transmit MRC

To obtain the channel information needed in transmit MRC, channel estimation should be carried out first. Pilot symbols could be sent from different transmit antennas in different time slots, so that the channel fading factor from the transmit antenna \( h_j \) to the receive antenna, \( h_{1,j} \), can be estimated. Then the receiver feeds the channel information \( h = (h_{1,1}, h_{1,2}, \cdots, h_{1,n_r}) \) back to the transmitter. When transmit MRC is employed, the weighting vector at the transmitter is given by

\[
\mathbf{w} = \mathbf{h}^{-1} \left( h_{1,1}^*, h_{1,2}^*, \cdots, h_{1,n_r}^* \right) = \mathbf{h}^{-1} \cdot \mathbf{h}^H
\]

(1.16)

where \( \|\mathbf{h}\|^{-1} \) is introduced to normalize the total transmit signal power to \( P \). So the transmit signal vector becomes \( \mathbf{w} \cdot \mathbf{s} \) and the received signal is given by

\[
y = \mathbf{h} \cdot \mathbf{w} \cdot \mathbf{s} + n = \|\mathbf{h}\|^{-1} \cdot \mathbf{h} \cdot \mathbf{h}^H \cdot \mathbf{s} + n = \|\mathbf{h}\| \cdot s + n
\]

(1.17)

whose SNR can be calculated as

\[
\text{SNR} = \frac{E\left[ \|\mathbf{h}\|^2 \right]}{E[\|\mathbf{n}\|^2]} = \frac{\|\mathbf{h}\|^2 \cdot P}{\sigma^2} = \frac{P}{\sigma^2} \sum_{j=1}^{n_r} |h_{1,j}|^2
\]

(1.18)

Therefore, a transmit diversity of order \( n_r \) can be achieved. It can be seen that when the number of transmit and receive antennas are the same, i.e., \( n_t = n_r \), the MISO system with transmit MRC can provide the diversity gain as that of the SIMO system with MRC.

1.5.2. Space Time Coding

Even if channel information is not available at the transmitter, MISO systems can still achieve transmit diversity by using space time coding. For example, an Alamouti code can be employed in a MISO system with two transmit antennas, as shown in Figure 1.2. The received signal vector \( \mathbf{y} \) is processed by a matched filter \( \mathbf{H}_a^H \), and the resultant signal is given by (1.4)

\[
\mathbf{z} = \left( |h_{1,1}|^2 + |h_{1,2}|^2 \right) \left( s_1 \right) + \left( h_{1,1}^* n_1 + h_{1,2}^* n_2 \right) + \left( h_{1,2}^* n_1 - h_{1,1}^* n_2 \right)
\]

To keep the total transmission power to \( P \), the signal power of \( s_1 \) and \( s_2 \) are set to \( E[|s_1|^2] = E[|s_2|^2] = P/2 \). It can be calculated that the SNRs for each element at the output of the STBC decoder are the same and given by...
\[ \text{SNR} = \frac{E \left\{ \left( |h_{1,1}|^2 + |h_{1,2}|^2 \right) s_i^2 \right\}}{E \left\{ |h_{1,1}^* n_1 + h_{1,2}^* n_2|^2 \right\}} = \frac{\left( |h_1|^2 + |h_2|^2 \right) \cdot P}{2\sigma^2} \]

(1.19)

Compared to the SNR obtained by transmit MRC (1.18), it can be seen that there are 3dB loss in the SNR provided by Alamouti code, due to the lack of channel information at the transmit side. Assuming two transmit antennas, QPSK modulation, and i.i.d. rayleigh fading channels, the bit error rate (BER) is illustrated in Figure 1.8 as a function of SNR when transmit MRC and Alamouti code are employed. Perfect channel information is available at both transmitter and receiver. As a comparison, the performance of system with single antenna is also shown. It can be seen that with more transmit antennas and proper signal processing technique, the system achieves better performance. Furthermore, the performance of transmit MRC is better than that of Alamouti code, at the cost of more complexity since channel information needs to be fed back to the transmitter for transmit MRC.

![Figure 1.8. Performance of a MISO system with two transmit antennas](image1)

![Figure 1.9. A MIMO system](image2)
1.6. MIMO

Figure 1.9 illustrates the structure of a MIMO system, where \(n_T\) and \(n_R\) antennas are deployed at the transmitter and receiver, respectively. MIMO can be designed to exploit the diversity gain. In this case, the same data symbol is transmitted on all \(n_T\) transmit antennas. The total diversity gain is not larger than \(n_T \cdot n_R\). MIMO can also be used for spatial multiplexing to increase the system throughput, when different data symbols are sent on different transmit antennas. In this case, the throughput can be increased linearly with the minimum number of transmit and receive antennas in the large SNR regime. There is always a tradeoff between the diversity gain and the throughput gain [25].

1.6.1. Background Knowledge

Define a matrix \(H\) of size \(n_R \times n_T\). It can be singular value decomposed (SVD) into three matrices as follows

\[
H = U_H \Sigma_H V_H^H
\]

(1.20)

where \(U_H\) and \(V_H\) are \(n_R \times n_R\) and \(n_T \times n_T\) unitary matrices constrained by \(U_H U_H^H = I\) and \(V_H V_H^H = I\), respectively, and \(\Sigma_H\) is a \(n_R \times n_T\) matrix with singular values of \(H\) as its diagonal elements. A non-negative real number \(\xi_i\) is a singular value for \(H\) if and only if there exists unit-length vectors \(u\) of size \(n_T\) and \(v\) of size \(n_R\), such that

\[
H v = \xi_i u \quad \text{and} \quad H^H u = \xi_i v.
\]

The vectors \(u\) and \(v\) are called left-singular and right-singular vectors for \(H\), respectively.

1.6.2. Dominant Eigenmode

When channel information is available at the transmitter, the dominant eigenmode algorithm can be employed to maximize the spatial diversity gain. In this scheme, the same data symbol \(s\) will be firstly weighted by a vector \(w_j = \left( w_{j,1}, \cdots, w_{j,n_T} \right)^T\), then the \(j\)th weighted copy of \(s\), \(w_j s\), is radiated from the transmit antenna \(#j\). Denote \(H\) as the MIMO channel matrix of size \(n_R \times n_T\), whose element \(h_{i,j}\) stands for the channel fading factor from the transmit antenna \(#j\) to the receive antenna \(#i\). Then, the receive signal vector is given by

\[
y = H w_j \cdot s + n\]

(1.21)

where \(n\) is a \(n_R \times 1\) noise vector with a covariance matrix of \(\sigma^2 I\). The received signals will be further weighted by a vector \(w_r = \left( w_{r,1}, \cdots, w_{r,n_R} \right)^T\) as follows

\[
z = w_r^H y = w_r^H H \cdot w_j \cdot s + w_r^H n\]

(1.22)

The SNR of \(z\) can be calculated as

\[
\text{SNR} = \frac{E \left( \|w_r^H H \cdot w_j \cdot s\|^2 \right)}{E \left( \|w_r^H n\|^2 \right)} = \frac{\left\|w_r^H H \cdot w_j \right\|^2}{\|w_r\|^2} \frac{P}{\sigma^2}
\]

(1.23)

Let \(\xi_{\text{max}} = \max \{\xi_1, \xi_2, \cdots, \xi_{\text{rank}(H)}\}\) represent the maximum singular value of \(H\), where \(\text{rank}(H)\) is the rank of \(H\). The receive SNR can be maximized when \(w_r\) and \(w_j\) are chosen as the left and right singular vectors of \(H\) corresponding to \(\xi_{\text{max}}\), respectively. The resultant SNR is given by \(\text{SNR} = \frac{\xi_{\text{max}}^2 P}{\sigma^2}\).
1.6.3. Space Time Coding

As in MISO systems, when there is no channel information at the transmitter, space time coding can be employed in MIMO systems to achieve diversity gains. Again, Alamouti code is taken as an example and a MIMO system with two transmit antennas and two receive antennas is considered. At the first and second time slot, the transmit signal vectors are \((s_1, s_2)^T\) and \((s_1', -s_2')^T\), respectively, which is the same as that in MISO systems. But there are two receive antennas now, and the corresponding received signal vectors are given by

\[
y_1 = \begin{pmatrix} h_{1,1} & h_{2,1} \\ h_{1,2} & h_{2,2} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \begin{pmatrix} n_{1,1} \\ n_{2,1} \end{pmatrix}
\]

and

\[
y_2 = \begin{pmatrix} h_{1,1} & h_{2,1} \\ h_{1,2} & h_{2,2} \end{pmatrix} \begin{pmatrix} s_1' \\ -s_2' \end{pmatrix} + \begin{pmatrix} n_{1,2} \\ n_{2,2} \end{pmatrix}.
\]

These signal vectors can be rearranged into a new vector as follows

\[
y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} h_{1,1} & h_{2,1} \\ h_{1,2} & h_{2,2} \end{pmatrix} \begin{pmatrix} s_1 \\ s_1' \end{pmatrix} + \begin{pmatrix} n_{1,1} \\ n_{1,2} \end{pmatrix} = H_{\text{eff}} s + n
\]

Similarly, \(y\) will pass through a matched filter of \(H_{\text{eff}}^H\). It can be proved that the equivalent channel matrix \(H_{\text{eff}}\) is orthogonal, i.e., \(H_{\text{eff}}^H H_{\text{eff}} = \|H_{\text{eff}}\|^2 I\). Therefore, the output of the matched filter is given by

\[
z = H_{\text{eff}}^H y = H_{\text{eff}}^H H_{\text{eff}} s + H_{\text{eff}}^H n = \|H_{\text{eff}}\|^2 s + H_{\text{eff}}^H n
\]

Thus, the receive SNR can be calculated as

\[
SNR = \frac{E\left[\left\|H_{\text{eff}} s\right\|^2\right]}{E\left[\left\|H_{\text{eff}} n\right\|^2\right]} = \frac{\left\|H_{\text{eff}}\right\|^4 P}{\left\|H_{\text{eff}}\right\|^4 \sigma^2} = \frac{\left\|H_{\text{eff}}\right\|^2 P}{2\sigma^2}
\]

Consider a 2 by 2 MIMO system with QPSK modulation and i.i.d. rayleigh fading channels. Perfect channel information is assumed at both the transmitter and receiver. The performance of the system using dominant eigenmode and Alamouti coding are shown in Figure 1.10. It can be seen that using dominant eigenmode, the system obtains a 3dB gain in SNR compared to that using Alamouti scheme.

1.6.4. Multiple Eigenmode Transmission

When channel information is available at the transmit side, the multiple eigenmode can be employed to maximize the throughput gain (or spatial multiplexing gain). As shown in Figure 1.11, \(n_r\) different data symbols \(s = (s_1, \ldots, s_n)^T\) are firstly weighted by a matrix \(W_t\), then radiated from different transmit antennas. After passing through the MIMO channel \(H\), the received signal vector is further weighted by another matrix \(W_r^H\). Assume that the SVD of \(H\) is given by \(H = U_H \Sigma_H V_H^H\). Then, \(V_H\) and \(U_H\) are chosen as the transmit and re-
receive weighting matrix, $W_t$ and $W_r$, respectively. At the receive side, the weighted output is given by

$$
z = W_t^H HW_t s + W_r^H n = U_H^H (U_H \Sigma_H U_H^H) V_H^H s + W_r^H n = \Sigma_H s + n' \tag{1.29}
$$

Figure 1.10. Performance of dominant eigenmode and Alamouti scheme

Figure 1.11. Multiple eigenmode transmission
Since $\Sigma_H$ is an $n_R \times n_T$ matrix with singular values of $H$ as its diagonal elements, $n_R$ must be equal to or larger than $n_T$ so that the transmitted signal vector $s$ can be recovered. Denote the $n_T$ singular values of $H$ as $\xi_1, \cdots, \xi_{n_T}$. The $i^{th}$ ($i \in [1, n_T]$) element of vector $z$ is expressed as

$$z_i = \xi_i s_i + n_i^* \quad (1.30)$$

It can be seen that using multiple eigenmode, the MIMO system with channel $H$ can be transformed into $n_T$ parallel SISO systems, as shown in Figure 1.11.

### 1.6.5. Signal Detection in MIMO Systems

When channel information is not known at the transmitter, but can be obtained by means of channel estimation at the receiver, multiple eigenmode is not applicable. Without pre-processing at the transmitter, the received signal becomes

$$y = H \cdot s + n \quad (1.31)$$

Various signal detection algorithms have been proposed to recover the transmitted signal vector $s$.

#### A. Linear Detection

Linear detection provides low-complexity solutions to the MIMO signal detection problem. Two popular linear detection algorithms are zero-forcing (ZF) (also known as least square (LS)) and MMSE detection. ZF detection is designed to minimize the square error between the received signal vector $y$ and the recovered transmitted signal vector $H \cdot s$, assuming that elements of $s$ take continuous values. The objective function is given by

$$s_{ZF} = \arg \min_{s = [s_1, \cdots, s_{n_T}]} \| y - H \cdot s \|^2 \text{ for continuously valued } s_1, \cdots, s_{n_T} \quad (1.32)$$

It will be shown later that if $s_1, \cdots, s_{n_T}$ take discrete values, the LS problem becomes a ML detection problem. By relaxing $s_1, \cdots, s_{n_T}$ to be continuously valued, the solution of the LS problem can be obtained by simply setting the derivation of $\| y - H \cdot s_{ZF} \|^2$ with respect to $s_{ZF}$ to zero

$$\frac{\partial \| y - H \cdot s_{ZF} \|^2}{\partial s_{ZF}} = 0 \Rightarrow H^H H \cdot s_{ZF} = H^H y \Rightarrow s_{ZF} = (H^H H)^{-1} H^H y = H^H y \quad (1.33)$$

where $H^H$ is the pseudo inverse of the $n_R$ by $n_T$ channel matrix $H$. To ensure that $H^H H$ is invertible, it is required that the rank of $H$ should be $n_T$. Therefore, to apply ZF detection, the number of receive antennas $n_R$ is required to be equal to or larger than the number of transmit antennas $n_T$, i.e., $n_R \geq n_T$.

Assuming a 2 by 2 MIMO system with QPSK modulation, i.i.d. rayleigh fading channels and practical channel estimation at the receiver [26], the performance of ZF detection is shown in Figure 1.12 with that of a SISO system. It can be seen that the BER of MIMO with ZF is the same as that of SISO. However, it should be noted that using the same spectrum, data information conveyed by the MIMO system is doubled compared to that of the SISO system. So ZF detection can only provide throughput gain but not diversity gains.

MMSE detection aims to design a weighting matrix $W_e$ to minimize the mean square error between the transmitted signal vector $s$ and the weighted received signal vector $W_e \cdot y$, given by

$$W_e = \arg \min_W E \left\{ \| s - W \cdot y \|^2 \right\} \quad (1.34)$$

Assume that the covariance matrix of signal and noise vectors are $E \{ s \cdot s^H \} = P/n_T \cdot I$ and $E \{ n \cdot n^H \} = \sigma^2 I$, respectively. It can be derived that
Figure 1.12. Performance of MIMO systems with ZF detection

Figure 1.13. Performance comparison between ZF and MMSE detection
Generally, MMSE detection can provide better performance than ZF detection. However, MMSE detection also needs more information than ZF, such as the channel noise variance and transmit signal power. Moreover, at high SNR, $\sigma^2 (P/n_r)^{-1}$ approaches zero and MMSE detection reduces to ZF detection. A performance comparison between the ZF and MMSE detection is given in Figure 1.13.

B. Nonlinear Detection: Interference Cancellation

MIMO systems can be taken as a special multi-user system. At the receive side, each antenna receives a combined signal containing data signals transmitted from different antennas. Consider the signal $s_i$ from the transmit antenna #i. Other data signals become interferences to $s_i$. The interference cancellation techniques can be applied in the signal detection of MIMO systems. The basic idea is to firstly use linear detections to recover the data signal from transmit antenna #n. Then the interference caused by $s_n$ can be cancelled out from the received signals. After interference cancellation, linear detections are used again to recover the data signal $s_{n-1}$. This linear detection and successive interference cancellation (SIC) process goes on until $s_1$ is obtained. When ZF or MMSE detection is applied with SIC, the scheme is known as ZF-SIC or MMSE-SIC. Assuming a 2 by 2 MIMO system, the ZF/MMSE-SIC algorithm is illustrated in Figure 1.14. Using QR decomposition, the ZF/MMSE-SIC algorithm can be realized in a simple way. Firstly, the QR decomposition of the channel matrix $H$ is given by $H = Q \cdot R$, where $Q$ is an $n_r \times n_r$ orthogonal matrix and $R$ is an $n_r \times n_r$ upper triangular matrix. After multiplying the received signal vector $y$ with the hermitian of $Q$, the result is given by

$$z = Q^\dagger y = Q^\dagger H \cdot s + Q^\dagger n = R \cdot s + Q^\dagger n$$ (1.36)

Letting $z_i$, $r_{i,j}$, $s_i$ and $n_i^*$ stand for the elements of $z$, $R$, $s$ and $Q^\dagger n$, respectively, the previous equation can be expressed in an elemental format as follows

$$W_r = \left( E \{ y y^H \} \right)^{-1} \cdot E \{ s \cdot y^H \} = E \{ s \cdot s^H \} H^H \left( H \cdot E \{ s \cdot s^H \} H^H + E \{ n \cdot n^H \} \right)^{-1}$$

$$= \left( H^H H + \sigma^2 (P/n_r)^{-1} \cdot I \right)^{-1} H^H$$ (1.35)
Using ZF criteria, $s_{n_r}$ can be simply recovered as $\hat{s}_{n_r} = \text{dec}\left(\frac{z_{n_r}}{r_{n_r,n_r}}\right)$, where $\text{dec}(\cdot)$ is the decision operation choosing the signal constellation that is the nearest to $z_{n_r}/r_{n_r,n_r}$. Then, the interference caused by $s_{n_r}$ can be cancelled out by subtracting $r_{1,n_r}\hat{s}_{n_r}, \ldots, r_{n_r-1,n_r}\hat{s}_{n_r}$ from $z_1, \ldots, z_{n_r-1}$, respectively. Next, $s_{n_r-1}$ can be recovered as $\hat{s}_{n_r-1} = \text{dec}\left(\frac{z_{n_r-1} - r_{n_r-1,n_r}\hat{s}_{n_r}}{r_{n_r,n_r}}\right)$, and so on. MMSE-SIC can also be implemented in a similar way.

Given a 2 by 2 MIMO system, the performance of ZF/MMSE-SIC is demonstrated in Figure 1.15 with comparison to ZF/MMSE detection. It can be seen that by using SIC based detection, the system performance is improved. Moreover, MMSE provides better performance than ZF, either with or without SIC.

It is well-known that in multi-user SIC, the signal with larger SNR should be recovered first. Similarly, in ZF/MMSE-SIC of MIMO systems, the signals should be detected in an appropriate order. Generally, the post-ZF/MMSE detection SNR is chosen as the criterion. That means, if $s_1, \ldots, s_{n_r}$ are recovered using ZF/MMSE detection, SNR can be calculated for each data decision variable at each time of detection. Suppose that $s_i$ has the highest SNR, then $s_i$ should be the first data signal to be recovered. After the interference caused by $s_i$ is cancelled out, the post-ZF/MMSE detection SNR can be calculated for the rest signals, and the one with the highest SNR should be the second signal to be recovered. At the cost of additional complexity caused by the ordering, the ordered ZF/MMSE-SIC schemes can provide better performance than the non-ordered ones.

\[
\begin{align*}
z_1 &= r_{1,1}s_1 + r_{1,2}s_2 + \cdots + r_{n_r,1}s_{n_r-1} + r_{n_r,n_r}s_{n_r} + n'_1 \\
z_2 &= r_{2,1}s_1 + \cdots + r_{n_r,2}s_{n_r-1} + r_{n_r,n_r}s_{n_r} + n'_2 \\
&\vdots \\
z_{n_r-1} &= r_{n_r-1,n_r-1}s_{n_r-1} + r_{n_r-1,n_r}s_{n_r} + n'_{n_r-1} \\
z_{n_r} &= r_{n_r,n_r}s_{n_r} + n'_{n_r} \\
\end{align*}
\]
The ordered ZF-SIC was firstly employed in Bell Labs Layered Space-Time (BLAST) or layered space time coding [14]. BLAST is a bandwidth-efficient approach to wireless communication which takes advantage of the spatial dimension by transmitting and detecting a number of independent co-channel data streams using multiple, essentially co-located, antennas. The central idea behind BLAST is the exploitation, rather than the mitigation, of multipath effects in order to achieve very high spectral efficiencies (bits/sec/Hz), when multipath is viewed as an adversary rather than an ally.

Figure 1.16 shows a typical system structure of a BLAST system. At the transmit side, the original symbol stream is divided into several sub-streams, each to be transmitted by one antenna. At the receiver, an array of antennas is used to pick up the multiple transmitted substreams and their scattered images. Under the widely used theoretical assumption of independent Rayleigh scattering, the theoretical capacity of the BLAST architecture grows roughly linearly with the number of antennas, even when the total transmitted power is held constant [27]. Depending on the mapping method between parallel symbol streams and transmit antennas, there are two schemes: V-BLAST, D-BLAST. The mapping method of V-BLAST is depicted in Figure 1.17 where each sub-stream will always be transmitted by the same antenna over the whole block, while the mapping method of D-BLAST is depicted in Figure 1.18 where each sub-stream will be transmitted from one antenna to another for each individual symbol.

The performance of V-BLAST codes in slow fading environments is inferior to that of D-BLAST codes. For V-BLAST, the average pairwise error probability of the bottommost row is inversely proportional to the \((n_R - n_T + 1)^{th}\) power of SNR. In contrast, in D-BLAST, the average pairwise diagonal error probability between two diagonals \(c\) and \(e\) is inversely proportional to the \((n_R - \text{sum}(\text{dis}(c,e)))^{th}\) power of SNR. Therefore, if
Figure 1.18. D-BLAST

constituent codes of equivalent data rate and complexity are deployed, the error probability of a D-BLAST code in a slow fading environment can be much lower than that of a V-BLAST code. In case of fast fading channel, they will perform almost the same.

C. Nonlinear Detection: Maximum Likelihood (ML) Detection

Among all signal detection algorithms, ML is the optimum detection providing the best performance at the cost of the highest complexity, which increases exponentially with the number of transmit antennas and the size of the signal constellation. The objective of ML is to find a signal vector \( \hat{s} \) that minimizes the Euclidean distance between \( y \) and \( H \cdot \hat{s} \)

\[
\hat{s} = \arg \min_{\hat{s}} \| y - H \cdot \hat{s} \|^2
\]  

(1.38)

Assuming that all transmitted data symbols are taken from the same signal constellation with size \( C \), the ML detection must search over \( C^n \) possible signal vectors to obtain \( \hat{s} \). Given a simple 2 by 2 MIMO system with QPSK modulation, the system performance with ML detection is illustrated in Figure 1.19. As a comparison, the performance of other detections is also shown. It can be seen that ZF detection presents the highest BER, while MMSE, ZF-SIC and MMSE-SIC provide better and better performance. Yet the BER curves of ZF, MMSE, ZF-SIC and MMSE-SIC have similar slope. Remember that ZF detection cannot provide diversity gain. Thus the diversity order of ZF, MMSE, ZF-SIC and MMSE-SIC are basically the same. But the BER of ML detection decreases much more rapidly compared to those of linear detections and interference cancellation schemes. Therefore, besides the throughput gain, ML detection can also provide a diversity gain.

Although ML detection can be implemented in simple systems, its complexity becomes unmanageable with increased number of antennas and high-level modulations. For example, using 64QAM modulation and four transmit antennas, there are totally \( C^n = 64^4 = 1.7 \times 10^{17} \) signal vectors to be searched. Due to its high complexity, ML is not applicable in practical systems. Some sub-optimum algorithms have been proposed, aiming to achieve near optimum performance with limited complexity.

1. QR-MLD

Given the QR decomposition of the channel matrix \( H \), i.e., \( H = Q \cdot R \), the objective function of ML can be expressed as [28]

\[
\begin{align*}
\hat{s} &= \arg \min_{\hat{s}} \| y - H \cdot \hat{s} \|^2 = \arg \min_{\hat{s}} \left( y - H \cdot \hat{s} \right)^H \left( y - H \cdot \hat{s} \right) \\
&= \arg \min_{\hat{s}} \left( Q^H y - Q^H H \cdot \hat{s} \right)^H \left( Q^H y - Q^H H \cdot \hat{s} \right) = \arg \min_{\hat{s}} \| Q^H y - R \cdot \hat{s} \|^2
\end{align*}
\]  

(1.39)

Letting \( y' \), \( r_{ij} \) and \( \hat{s} \) stand for the elements of \( Q^H y \), \( R \), and \( \hat{s} \), respectively, the objective function can be further written as
This expression indicates that ML algorithm can be realized in a tree structure, as shown in Figure 1.20.
from the root of $\hat{s}_{n_r}$. Since $\hat{s}_{n_r}$ has $C$ different values, $C$ branches can be generated with $d_{n_r} = \left| y_{n_r}^t - \left( r_{n_r, n_s} \hat{s}_{n_s} \right) \right|^2$ as the weight on each branch. So there are $C$ candidates of $\hat{s}_{n_r}$ at the first layer. Then, for each branch of $\hat{s}_{n_r}$, $C$ sub-branches are generated according to the values of $\hat{s}_{n_r-1}$ with $d_{n_r-1} + d_{n_r}$ as the sub-branch weight. As a result, $C^2$ candidates of $\{\hat{s}_{n_r-1}, \hat{s}_{n_r}\}$ are obtained at the second layer. Finally, at the $n_r^{th}$ layer, there will be $C^{n_r}$ candidates of $\{\hat{s}_1, \cdots, \hat{s}_{n_r}\}$ with $d_1 + \cdots + d_{n_r}$ as the branch weight. The ML solution is simply the sequence $\{\hat{s}_1, \cdots, \hat{s}_{n_r}\}$ with the minimum branch weight.

To reduce the complexity of ML detection, the number of candidates at each layer can be decreased. For example, at the first layer, instead of keeping all $C$ branches/candidates, only the $C_1$ branches with smaller weight $d_{n_r}$ are kept and the other branches are deleted. Therefore, at the second layer, there will be $C_1 \cdot C$ branches. Similarly, only $C_2$ out of $C_1 \cdot C$ branches are kept for the next layer. Assuming $C_0 = 1$, it can be shown that the total number of branches visited by this sub-optimum ML detection is $C + C_1 \cdot C + C_2 \cdot C + \cdots + C_{n_r-1} \cdot C = C \cdot \sum_{n=0}^{n_r-1} C_n$ , while for full ML detection, the number will be $C + C^2 + \cdots + C^{n_r} = \left( C^{n_r+1} - C \right) / (C-1)$. By choosing different values for $C_n$, the sub-optimum algorithm can provide various system performance with different levels of complexity.

II. Sphere Decoding

The sphere decoding algorithm adds a condition to the objective function of ML detection, i.e.,

$$\left| \mathbf{Q}^H \mathbf{y} - \mathbf{R} \cdot \hat{s} \right|^2 \leq r_d^2,$$

where $r_d$ is a preset value [29]. Since $\mathbf{Q}^H \mathbf{y}$ is a point in a $n_1$-dimensional vector space, the conditional ML problem is equivalent to find a point $\mathbf{R} \cdot \hat{s}$ in a sphere of radius $r_d$ centered at $\mathbf{Q}^H \mathbf{y}$ that has the minimum distance to $\mathbf{Q}^H \mathbf{y}$. Since the searching is limited in the sphere, the number of vectors that should be visited is reduced compared to that of the full ML algorithm. Thus, the complexity is decreased.

Based on the tree structure, the sphere decoding algorithm can be described as follows. Similar to QR-MLD, the algorithm begins with the element $\hat{s}_{n_r}$. Firstly, find all possible values of $\hat{s}_{n_r}$ that satisfy

$$\left| y_{n_r}^t - \left( r_{n_r, n_s} \hat{s}_{n_s} \right) \right|^2 \leq r_d^2.$$  \hspace{1cm} (1.41)

Then, choose one value of $\hat{s}_{n_r}$ and find all candidates of $\hat{s}_{n_r-1}$ that satisfy the following expression

$$\left| y_{n_r-1}^t - \left( r_{n_r-1, n_s} \hat{s}_{n_s} + r_{n_r-1, n_r} \hat{s}_{n_r} \right) \right|^2 \leq r_d^2.$$  \hspace{1cm} (1.42)

If there is no such candidate of $\hat{s}_{n_r-1}$, choose another value of $\hat{s}_{n_r}$ until the solution set of the inequality is not null. Next, given $\{\hat{s}_{n_r-1}, \hat{s}_{n_r}\}$, all possible values of $\hat{s}_{n_r-2}$ can be obtained by solving the inequality

$$\sum_{n=n_r-2}^{n_r-1} \left| y_{n_r}^t - \sum_{n=n_r}^{n_r-1} r_{n_r, n_s} \hat{s}_{n_s} \right|^2 \leq r_d^2.$$  \hspace{1cm} (1.43)

This searching goes on until the candidates of $\hat{s}_1$ are generated given $\{\hat{s}_1, \cdots, \hat{s}_{n_r}\}$, providing solutions $\hat{s} = \left( \hat{s}_1, \cdots, \hat{s}_{n_r} \right)$ to the inequality $\left| \mathbf{Q}^H \mathbf{y} - \mathbf{R} \cdot \hat{s} \right|^2 \leq r_d^2$. Calculate the distances between $\mathbf{R} \cdot \hat{s}'$ and $\mathbf{Q}^H \mathbf{y}$, and choose the one with the minimum distance $\hat{s}'$. This is not the final solution since only
one possible candidate of \( \{ \hat{s}_1, \ldots, \hat{s}_m \} \) is considered. But this solution can be used to update the radius \( r_u \) since the distance between \( R \cdot \hat{s}^* \) and \( Q^H y \) should be less than the initial radius with high probability. Then the searching goes on in a smaller sphere with the updated radius \( r_u \) until all points in the sphere are visited.

In sphere decoding, it is important to set an appropriate value for the initial radius. If the initial radius is too small, there may be no points in the sphere and the solution of the conditional ML problem is null. On the other hand, if a large initial radius is set to ensure that the ML solution is included in the sphere, there will be many points in the sphere and the complexity is high. Generally, the initial radius can be set according to the quality of the channel, i.e., the noise variance. When the noise is large, the channel is in bad condition and it is likely that the ML solution is far from the center of the sphere. Thus, a large radius is needed. On the contrary, if there is little noise, the ML solution should be close to the center of the sphere and a small radius should be chosen to reduce the complexity. Another way to set the initial radius is that a simple linear detection is carried out first to find a candidate of transmitted signal vector. Then the distance between the solution of linear detection and the received signal vector can be taken as the radius. This method ensures to include the ML solution in the sphere.

Using the tree structure of ML problem, it can be seen that in QR-MLD, at each layer of the tree, some branches are deleted based on the current weights. Since the deleting is not based on the whole sequence of \( \{ \hat{s}_1, \ldots, \hat{s}_m \} \), it is possible that the ML solution is excluded as some branches are removed. Thus, the solution of the QR-MLD may not be a ML solution. But the complexity of QR-MLD is totally under control by choosing \( C_1, \ldots, C_{n-1} \). On the other hand, in sphere decoding, conditioned on the radius, each searching moves along the branches to the end so the whole sequence of \( \{ \hat{s}_1, \ldots, \hat{s}_m \} \) is considered. The radius is reduced based on the distance between \( R \cdot \hat{s}^* \) and \( Q^H y \). So sphere decoding can always provide the ML solution unless there is no solution due to small radius. The disadvantage of sphere decoding is that the complexity of the algorithm is a random variable since the number of points to be visited in a sphere is unknown. In some cases, sphere decoding need similar amount of computation as that of full ML detection, while sometimes the complexity of sphere decoding is near to that of linear detections.

### 1.7. Multi-User MIMO

All the MIMO systems described before are for single user, i.e., there is only one transmitter and one receiver. However, multi-user topology is even more popular in current communication systems where many mobile users communicate with one base station. There are two typical multi-user topologies: one is multiple access channel (MACH) which is corresponding to uplink communication, the other is broadcast channel (BCH) which is corresponding to downlink communication. For MACH, the system model can be expressed as

\[
y = (H_1 \ H_2 \cdots H_M) \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_M \end{bmatrix} + n
\]  

(1.42)

where \( y \) denotes the received signal vector at the BS, \( M \) is the number of users, \( s_i \) stands for the transmitted signal vector from the \( i \)th MS, and \( H_i \) is the channel matrix from the \( i \)th MS to the BS.

On the other hand, in case of BCH, the system model is given by

\[
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_M \end{bmatrix} \begin{bmatrix} s \\ n_1 \\ n_2 \\ \vdots \\ n_M \end{bmatrix}
\]  

(1.43)
where \( y_i \) denotes the received signal vector at the \( i \)-th MS, \( s \) is the transmitted signal vector from the BS and \( H_i \) is the channel matrix from the BS to the \( i \)-th MS. This model is valid for any number of antennas at the MS.

Finding the multi-user channel capacity has always been a hot topic [30,31]. Early researches focused on the achievable data rates for a vector Gaussian MACH where only the BS has multiple antennas. It has been shown that by using an optimum decision feedback multiuser equalizer and successive detection at the BS, the sum-rate of all users can achieve the maximal capacity of the channel [30]. On the other hand, the concept of water-filling can be employed at the MS side to achieve the system capacity [32,33,34]. Later, these works were extended to the case where multiple antennas are deployed at both BS and MS. By generalizing the aforementioned decision feedback equalizer to the vector access signal case [35], the achievable sum capacity is given by

\[
\sum_{i=1}^{M} R_i = \log \det \left( I + \sum_{i=1}^{M} H_i Q_i H_i^H \right)
\]

(1.44)

where \( Q_i \) is the covariance matrix of \( s_i \), i.e., \( Q_i = E \{ s_i s_i^H \} \). By designing the set of covariance matrix and power allocation schemes, the sum capacity (1.44) can be maximized. If each user has a fixed power constraint, an iterative water-filling method can be employed to solve the optimization problem [36]. The capacity can also be optimized under the overall power constraints for all users [37]. Although there has been much effort to evaluate the capacity region of the vector MACH, less is known on the design of actual coding/signal processing systems that take full benefit of the rich capacity of the channel. A set of transmit and receive vectors can be computed by using a direct SVD (D-SVD) method to exploit the spatial resource to enhance the system performance [38]. With the duality between uplink and downlink, the D-SVD method can also be applied in downlink scenario by interchanging the corresponding transmit and receive vectors of each user.

Comparing to vector MACH (uplink), the capacity region of vector BCH (downlink) is even more difficult to obtain and is still unknown to date. Due to the fact that BCH and MACH are dual to each other [39], Costa pre-coding method [40] in uplink can be used in downlink as well. The pre-coding method was later re-studied for the more general case where each MS has multiple antennas [36,41,42] and such studies have opened up a new branch of information theory called dirty-paper coding (DPC) [43]. However, the theory seems to be far from being useful for the design of efficient wireless communications.

Recent works on improving downlink capacity focus on beamforming approaches for realizing space division multiplexing [44,45]. Beamforming algorithms can be designed to maintain the signal to interference-plus-noise ratio (SINR) of every user at a preset value for acceptable signal reception [31,44]. The BS antenna array can also be operated using a simple maximal ratio transmission for diversity reinforcement at the MS [46]. Further performance improvement is obtained by using BS diversity array in combination with joint detection at the MS side [47]. There exists a closed-form antenna solution that optimizes the BS antenna array in maximizing the product of multi-user SINR [45]. In these approaches, the co-channel users are not truly uncoupled. The residual co-channel interference (CCI) will degrade users’ performance, and, most importantly, destroy the independency for managing multi-user signals, since the power of co-channel users must be carefully adjusted jointly. When dealing with multi-user communications, it is always advantageous to handle users in an orthogonal manner, as in conventional systems such as time, frequency or code division multiplexing (T/F/CDM) systems.

Orthogonal space division multiplexing (OSDM) in BCH can be achieved by using multi-user MIMO antenna. The spatial orthogonalization is realized by projecting every user’s signal onto the nullspace of the channels from the BS to the antennas of all the unintended mobile receivers. A set of transmit and/or receive weight vectors are used to separate different groups of streams among different users and (if possible) different streams of each user. By doing so, the multi-user system can be simplified to several independent single-user systems each with corresponding instant channel gains (similar to Figure 1.11), so that the FEC can be used for each user independently. Although the separation of spatial domain processing and time domain processing may degrade the overall system performance, this concept can simplify the system design a lot. A simple example of OSDM is to make the BCH block diagonal so that co-channel users are not interfered with each other [48,49,50].

In order to obtain the rich diversity of the channels and reduce the number of antennas at the BS, the OSDM optimization and receiver diversity combining should be considered jointly. Assuming that every MS uses a MRC receiver and the number of BS antennas is equal to the number of total co-channel signals co-existing in
the system, an iterative method can be employed to jointly optimize the BS and MS antenna weights for downlink OSDM to obtain diversity gains [45]. This method is finally extended to a more general case by Pan [51,52] as demonstrated in Figure 1.21 where each user can further support multiple independent streams.

1.8. Conclusions

Multiple antenna techniques are extensively reviewed in this chapter. For single-user point-to-point systems, space-time coding, SIMO and MISO schemes are introduced to provide spatial diversity. Moreover, precoding and combining is a simple way to exploit multiplexing gain when CSI is known at the transmitter. On the other hand, if channel state is only known at the receiver, efficient detection algorithms are needed, such as non-ordered ZF/MMSE-SIC, BLAST, ML detection, and sphere decoding. Finally, MIMO techniques are also discussed in multi-user scenarios.

References


About the Authors

Dr. Yiqing ZHOU received the B.S. degree in communications and information engineering and the M.S. degree in signal and information processing from the Southeast University, Nanjing, China, in 1997 and 2000, respectively. In February 2004, she received the Ph.D. degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong. From June 2004 to Oct. 2008, she was with the Department of Electrical and Electronic Engineering at the University of Hong Kong as a Postdoctoral Fellow. Since Nov. 2008, Dr. Zhou has been with Hong Kong Applied Science and Technology Institute Corp. as a principal engineer. She is leading a project of core IP R&D on LTE-Advanced. Dr. Zhou has published more than 30 papers in IEEE journals and international conferences. Her research interests include coding theory, spread spectrum, OFDM systems, interference cancellation, hybrid ARQ, MIMO and other wireless techniques for high speed data communications.

Dr. Zhengang PAN received the Bachelor and Master degrees, all in Radio Engineering, from Southeast University, Nanjing, China, in July 1997 and March 2000, respectively. During the year from Sept. 1997 to Dec. 2000, he was working in Nanjing Xuji Communication and Automation Co. Ltd. on the development of power-line communication equipment. From January 2001 to April 2004, he has been in Department of Electrical and Electronic Engineering, the University of Hong Kong, pursuing Ph.D degree and successfully obtained the degree in December 2004. After that, he joined DoCoMo Beijing Communication Labs Co. Ltd, working on the front-end research for the next generation wireless communication standards, including 802.11n, 802.16d/e, HSPA and LTE. He has involved in many technical fields including time/frequency/sampling synchronization technology for single-carrier/multi-carrier(OFDM)/A based system, channel estimation, forward error correction coding, multiple antennas systems (MIMO) and space-time processing/coding, cross layer optimization and so on. In the year 2006, he joined ASTRI, as one of the foundation team member of the Practical MIMO Core project. In this project, he
led a team to pursue the target as how to achieved the wireless channel capacity promised by MIMO theory under practical limitations such as small form-factors of the mobile terminal, mutual coupling between antenna elements, large power consumption by multiple RF chains and so on.

Dr. Kai-Kit WONG received the BEng, the MPhil, and the PhD degrees, all in Electrical and Electronic Engineering, from the Hong Kong University of Science and Technology, Hong Kong, in 1996, 1998, and 2001, respectively. After graduation, he joined the Department of Electrical and Electronic Engineering, the University of Hong Kong as a Research Assistant Professor. From July 2003 to December 2003, he visited the Wireless Communications Research Department of Lucent Technologies, Bell-Labs, Holmdel, NJ, U.S., where he was a Visiting Research Scholar studying optimization in broadcast MIMO channels. After that, he then joined the Smart Antennas Research Group of Stanford University as a Visiting Assistant Professor conducting research on overloaded MIMO signal processing. From 2005 to August 2006, he was with the Department of Engineering, the University of Hull, U.K., as a Communications Lecturer. Since August 2006, he has been with University College London Adastral Park Campus and currently moved to the Department of Electrical and Electronic Engineering where he is a Senior Lecturer. Dr Wong won the IEEE Vehicular Technology Society Japan Chapter Award of the International IEEE Vehicular Technology Conference-Spring in 2000, and was also a co-recipient of the First Prize Paper Award in the IEEE Signal Processing Society Postgraduate Forum Hong Kong Chapter in 2004. In 2002 and 2003, he received, respectively, the SY King Fellowships and the WS Leung Fellowships from the University of Hong Kong. Also, he was awarded the Competitive Earmarked Research Grant Merit and Incentive Awards in 2003-2004.Dr Wong is a Senior Member of IEEE and is also on the editorial board of IEEE Transactions on Wireless Communications, IEEE Communications Letters, IEEE Signal Processing Letters, and IET Communications.